

## Matching process

Matching is a probabilistic record linkage implementation. This process assigns numeric scores called weights to the comparison of individual data elements. The scores measure the contribution of each data element to the overall or composite weight. The composite weight, in turn, is the sum total weight of all defined comparisons. Some data elements contribute more weight because they are more critical to the match or more reliable than others. Statistical properties of the data elements and tuning parameters determine the weight of the contributions.

The weight for a data element is generated by using one of the comparison functions that are available in matching. Among the comparison functions are exact functions and functions that provide a full spectrum of error-tolerant or fuzzy matching functions. You can adjust the resulting weight of a given comparison function to reflect the importance of the data element to its domain and to the overall comparison.

The composite weight is compared against a set of thresholds (also called cutoffs) to determine how to translate the weight into a measurement of confidence. The confidence indicates the likelihood that a matching pair of records was identified.

### Matching process internal columns

This is the list of additional columns generated during matching process.

Column name	Description
<b>qsMatchSetID</b>	The match set identifier.
<b>qsMatchType</b>	Match type
<b>qsMatchDataID</b>	The match ID for the record. One of: MP, DA, CP, RA.
<b>qsMatchWeight</b>	The weight of the record.
<b>qsMatchPassNumber</b>	The number of the pass where the match was found.

### Matching record types

The Matching process output contains four types of records, which is informed in the generated column *qsMatchType*:

Record type	Description
<b>Match (MP)</b>	Master record
<b>Clerical (CP)</b>	The duplicates that fall in the clerical range.
<b>Duplicate (DA)</b>	The duplicate records that are above the match cutoff.
<b>Residuals or nonmatched (RA)</b>	The records that are not master, duplicate, or clerical records.

### Clerical/Match cutoff

Match and clerical cutoffs are thresholds that determine how to categorize scored record pairs. The goal of setting cutoffs is to minimize uncertainty in the match results while you limit the number of false categorizations.

Record pairs with composite weights equal to or greater than the match cutoff are considered matches. Record pairs with composite weights equal to or greater than the clerical cutoff but less than the match cutoff are called clerical pairs. The matching process is uncertain whether clerical pairs are matches or nonmatches. Pairs with composite weights below the clerical cutoff are considered nonmatches. You can set cutoffs at the same value, so that you eliminate clerical records.

You can set a high cutoff threshold to limit the results to better quality matches, though possibly fewer matches. A lower threshold can produce more matches, but some of these matches might be of lesser quality. Business requirements help drive decisions. Results can vary depending on whether you take a conservative or more aggressive approach to defining the cutoff values.

The following table shows the cutoffs for each pass:

Pass	Match cutoff	Clerical cutoff
1	1	1
2	1	1
3	20	7
4	3	1
5	2	0
6	0	0
7	12	0
8	36	0
9	30	0
10	26	0
11	20	0
12	50	50
13	1	1